(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 781 068 A1

# (12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
**25.06.1997 Bulletin 1997/26**

(51) Int. Cl.[6]: **H04Q 11/04**, H04L 12/56

(21) Application number: **95480182.5**

(22) Date of filing: **20.12.1995**

(84) Designated Contracting States:
**DE FR GB**

(71) Applicant: **INTERNATIONAL BUSINESS MACHINES CORPORATION**
**Armonk, NY 10504 (US)**

(72) Inventors:
- **Fichou, Aline**
  **F-06480 La Colle sur Loup (FR)**
- **Galand, Claude**
  **F-06800 Cagnes sur Mer (FR)**
- **Foriel, Pierre-André**
  **F-06700 Saint Laurent du Var (FR)**

(74) Representative: **Schuffenecker, Thierry**
**Compagnie IBM France,**
**Département de Propriété Intellectuelle**
**06610 La Gaude (FR)**

## (54) Method and system for adaptive bandwidth allocation in a high speed data network

(57) This adaptive bandwidth allocation for Non-Reserved traffic over high speed transmission links of a digital network is operated through regulation of data packet transfers over network nodes/ports including input/output adapters connected through a switching device.

To that end the network node is assigned with a Control Point computing devise (CP) storing a Topology Data Base keeping an image of the network.

This Data Base is periodically and at call set up updated by Topology Data Base Update messages (TDUs) including an Explicit Rate parameter for link l indicating the current available bandwidth on link l, and a parameter $N_{NRI}$ indicating the number of Non-Reserved connections on link l.

These informations are used within each Adapter to periodically regulate the transmission bandwidth assigned to each Non-Reserved traffic connection within the network. To that end, each adapter is provided with an Access Control Function device for each attached connection (data source) and a Connection Agent (CA) getting, on request, required current link informations from the attached Topology Data Base.

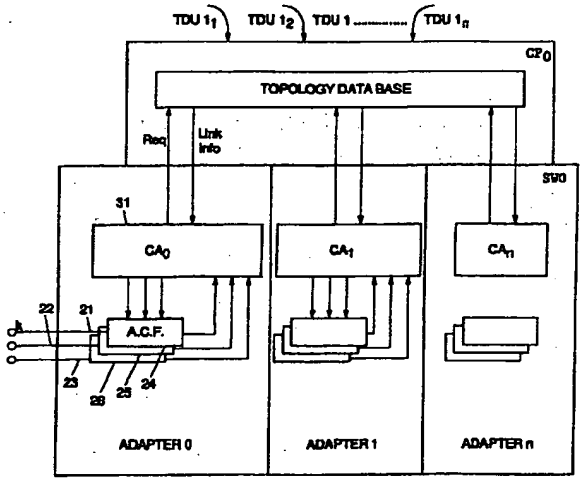


FIG.2

EP 0 781 068 A1

## Description

### Field of the Invention

This invention relates to a method and system for adaptively and dynamically operating bandwith allocation to so-called Non-Reserved Bandwidth traffic in a high speed-data transmission network.

### Background of the Invention

Modern digital networks are made to operate in a multimedia environment for transporting different types of data (pure data or digitized information signals including voice, image, video, etc) over the same network, while ensuring the compliance with the requirements specific to each kind of these traffics.

For instance, one may notice that the information provided by various users can be divided into different types. These include real-time information (such as voice information) that must be transmitted to the end-user with predefined limited time-delay restrictions, and non-real-time information. If some real-time information is not transferred within said time delay, it should simply be discarded.

In this case, recovery of the original signal, at end-user location is made possible to some extent, by providing techniques such as interpolation/extrapolation techniques, in a data packet transmission environment. These techniques do provide solutions to overcome the "loss" of only a limited number of discarded consecutive packets. They do not overcome the delay restriction requirement.

On one hand the information may be considered as including : a so-called Reserved traffic information whose transmission must be guaranteed with predefined maximum delay according to conditions contractually agreed upon by both parties, i. e. end-user and network owner ; and a Non-Reserved (NR) information, such as certain control information or traffic of specific sources to be vehiculated over the network with no specific time constraint, but which traffic should be optimized for network efficiency.

On the other hand, one should recall that different techniques have been developed, such as packet switching techniques, whereby the digitized data are arranged into so called bit packets as already mentionned, and circuit switching techniques.

The basic advantage of packet switching techniques as compared to circuit switching techniques, is to allow a statistical multiplexing of the different types of data over a line which optimizes the transmission bandwith. The drawback of packet switching techniques compared to circuit switching techniques, is that it introduces jitter and delay, which, as already considered, may be detrimental for the transmission of isochronous data, like video or voice. This is why methods have been proposed to control the network in such a way that delays and jitters are bounded for every new connection that is set-up across a packet switched network.

These methods have been described, for instance a co-pending European Application 94480097.8 . All these methods include, for any end user requesting service or any control data to be vehiculated over the network, establishing a path through network high speed links (or lines) and nodes or ports, with optimal use of the available transmission bandwidth.

Basically, one may allocate predefined bandwidths to Reserved traffic (including real-time) on the basis of contractually defined parameters, and then allocate whatever left in the bandwidth to Non-Reserved traffic on fixed basis.

But, bearing in mind that instantaneous traffics are eminently variable for both Reserved and Non-Reserved traffics, any fixed bandwidth assignment is naturally inefficient as far as the efficiency of global network utilization is concerned.

A first improvement consists in precomputed oversized bandwidths allocations on a source by source basis, with the provision of means for detecting instantaneous congestions occurring within the network and for monitoring some sort of "slowing-down" mechanism. Such a mechanism has already been described in the above mentioned copending European Application as well as in US patent 5 280 470. In the mentioned co-pending application, the slowing-down mechanism is only used to control congestion at node switching level through management of a switch back-pressure signal. In the US patent 5 280 470 slowing-down is operated over reserved bandwidths only when congestion is detected in network nodes and no increase is allowed, which may result in low link utilization in case of some sources being non-active. Accordingly, the data sources are not really taken into account on a dynamic basis. In other words, the considered prior art system does not provide true dynamic sharing among connections but only helps resolving traffic congestions temporarily by sending a slow-down message back to sources.

### Objects of the Invention

One object of this invention is to provide a method for optimizing bandwidth allocation to Non-Reserved traffic in a data communication network.

Another object of the invention is to provide a method for monitoring Reserved traffic and dynamically assigning or

adapting transmission bandwidth to Non-Reserved traffic accordingly in a data communication network.

Still another object of the invention is to provide a method for dynamically and fairly distributing transmission bandwidth among Non-Reserved traffic sources, based on real data sources requirements, in a communication network operating in Asynchronous Transfer Mode (ATM), or Frame relay.

These and other objects, characteristics and advantages of the invention, shall be more readily apparent from the following description of a preferred embodiment made with reference to the accompanying drawings.

## Summary of the Invention

This invention deals with an adaptive bandwidth allocation method for optimizing data traffic in a high speed data transmission network including Nodes interconnected by high speed links made for vehiculating data traffics of different priority levels along assigned network paths between data terminal equipments acting as data sources and data destination terminals, said priority levels including high priority level(s) for so-called Reserved traffic for which a transmission bandwidth has been reserved along said assigned path based on predefined agreements, and low priority level for so-called Non-Reserved traffic which should be transferred over the network within the transmission bandwidth available along the considered path once Reserved traffic is satisfied, said method for assigning bandwidth to Non-Reserved traffic including :

- generating and keeping at least one Topology Data Base storing an image of the network occupancy over each link along the network paths;
- periodically and at any terminal call set-up within the network, generating and broadcasting Topology Data Base updating (TDUs) messages, said TDU messages including a so-called Explicit Rate parameter ($ER_l$) for each link l indicating the bandwidth currently available on link l, and a parameter $N_{NR}$ indicating the number of Non-Reserved connections on link l;
- upon receving said TDUs information, computing for each node along the considered path the amount of transmission bandwidth left available over each link along the considered path and assigning said available bandwidth to the Non-Reserved traffic sources connected to the network.

## Brief Description of the Figures

Figure 1 is a schematic representation of an example of data transmission network made to include the invention.

Figure 2 is a schematic representation of the device made to implement the invention within one network node.

Figure 3 is a detailed representation of the system made to implement the invention within one node adapter.

Figure 4 is a schematic representation of a device to be used within the invention.

Figure 5 is a detailed flow-chart made to implement the invention.

## Detailed Description of a Preferred Embodiment of the Invention

While the invention applies equally to both centralized control networks and distributed control networks, the preferred embodiment shall herein be described with reference to a distributed control data transmission network. Therefore, in no instances should this be considered as implying any limitation to the invention.

Accordingly, Figure 1 shows an example of a packet switching network with distributed control, (i. e. control at each node of the network), which can be used to implement the invention. This network includes five switching nodes SW0 to SW4 interconnected by high speed trunk lines (or links) (10, 11, 12, 13 et 14) and that can be accessed by access lines (AL's) connected to outside data terminal equipments (DTE's) acting as data sources or destination terminals.

The network control architecture being a distributed one, each switching node is controlled by a control point CP. All CP's are interconnected via a control point spanning tree (CPST) which provides an efficient means for multicasting control messages between the control points. When a CP wants to broadcast a control message to other CP's in the network, it sends this message to the predefined CPST switching address, and the architecture provides the means to route this message on every line of the CP spanning tree, and only on these lines. This architecture also provides the means to initialize the tree address on each switching node, and to automatically reconfigure the tree in case of line or switching node failure.

Each CP includes a copy of the Topology Data Base that contains the information about the network. It includes the

network physical configuration and the line characteristics and statuses.

For every line l, in the network, vehiculating a so-called Reserved traffic, the maximum delay T(n) that can be accepted on a packet with a specified priority, and the level of bandwidth utilization R_res(n) on this line are defined and recorded in the topology data base. This information is distributed to other control points via topology data base update messages ($TDU_s$) sent over the control point spanning tree whenever required.

For more informations on such a Spanning Tree organization, one may refer to the Copending European Patent Application number 94480048.1, with the title "A Data Communication Network and Method for Operating said Network".

In operation, any source user terminal equipment may request being connected to a destination terminal. For instance user terminal equipment DTE-A and DTE-B which are respectively connected to the network via access lines AL-A and AL-B, shall be interconnected thru the network with a given quality of service (QoS) specified in terms of a maximum delay T-max and a packet loss probability P-loss, upon DTE-A requesting being connected to DTE-B (i.e. at DTE-A call set-up) .

To that end, at the switching node SW0, the control point CP0 first uses the QoS and the traffic characteristics specified by the user (peak rate, mean rate, average packet length) to compute the amount of bandwidth C_eq, called the equivalent capacity of the connection to be reserved on every line, on the route or path assigned to the traffic between source terminal and destination terminal, in order to guarantee a packet loss probability P0(n) on this line which is smaller than the loss probability P_loss that has been specified for the connection.

Based of the information that is available on a line basis in the topology data base, the control point CP0 then computes the best route in the network to reach the destination. To that end, a Path Selection Program first identifies the network lines that are eligible for the route. If R(n) and R_res(n) respectively denote the capacity of line n and its current level of reservation, then the line is eligible if :

$$R\_res(n) + C\_eq \leq 0.85\ R(n)$$

it then uses a modified Bellman-Ford algorithm to find the minimum weight, minimum hop count, route from the origin to the destination which uses eligible lines and which satisfies the QoS.

$$T\_max \leq \Sigma T(n)$$

$$P\_loss \leq 1-\pi(1-P\_l(n))$$

where the summation and product operators are carried over the N lines of the route (n=1, ..., N).

For additional information on equivalent capacity and best route considerations, one may refer to the following publications :

- R. Guerin, H. Ahmadi, M. Naghshineh, "Equivalent Capacity and its Application to Bandwidth Allocation in High Speed Networks", published in IEEE Journal of Selected Areas in Communications, JSAC-7, Sept. 1991.

- H. Ahmadi, J. S. Chen, R. Guérin, L. Gün, A. M. Lee and T. Tedijanto, "Dynamic Routing and Call Control in High-Speed Integrated Network", pusblished in Proc. Workshop Sys. Eng. Traf. Eng., ITC 13, pp 397-403, Copenhagen, Denmark.

Now, let's assume that the chosen route to connect DTE-A1 to DTE-B from switching node SW0 to switching node SW4 uses trunk lines 10, 11 and 14 via switching nodes SW1 and SW2 and access lines AL1 on switching nodes SW0 and SW4.

To that end, the origin control point CP0 sends a connection set-up message along the route, a copy of which is delivered to the control point of every switch on the route (e. g. CP1, CP2 and CP4). This message contains a list of the network addresses of the control points on the route, the list of the link names (e. g. 10, 11, 14) between these control points, the request bandwidth C_eq, the priority of the connection, and a connection correlator C_cor which is set by the origin control point CP0 and which is used by all other CP's to uniquely identify the connection.

Upon reception of the copy of the set-up message, each CP performs two basic tasks.

First, the CP checks whether the equivalent capacity of the new connection is still available on the line to the next switching node on the route, and if it is available, it reserves it. Therefore, the CP checks whether the line is eligible by verifying the first above relation. If it is, the CP reserves the desired amount of bandwidth on the transmit line for the new connection, accepts the connection set-up, increments the reservation level :

$$R\_res(n) = R\_res(n) + C\_eq$$

and if this reservation level has significantly changed, it eventually broadcasts a topology data update (TDU) message on the CP spanning tree to inform the other CP's of the new reservation level R_res of this particular line.

Second, the CP allocates a new label for the new connnection, and sends back this label to the control point of the previous switch on the route, for label swapping purposes.

Accordingly, during network operation, each node n (n=0, 1,...) Control Point (CPn) periodically, and at call set-ups, broadcasts Topology Data-Bases Updating (TDUs) messages collected by access nodes. Now, for the purpose of the dynamically adaptive control of the network bandwidth allocation of this invention, the TDU format shall be made to include a so-called Explicit Rate (ER) parameter. Accordingly, as represented in Figure 1, the TDU format for any link l shall include, in the above mentioned TDU messages, the Explicit Rate for link l (i. e. ERl) which ERl shall specify the current available bandwidth on the link l divided by the number of NR connections, and therefore indicate dynamically said available bandwidth to the corresponding Control Points. This information shall enable adaptively assigning bandwidth to sources requesting service for transmitting the so-called Non-Reserved (NR) traffic, with a full knowledge of the bandwidth currently available on all considered links.

Assuming the traffic to be vehiculated from source terminal DTE-A1 to destination terminal DTE-B, through the above mentioned path SW0, 10, SW1, 11, SW2, 14, SW4, then each of these node Topology Data Bases shall have available the Explicit Rates on the node connected trunks.

In addition, the TDU format over link l shall also include an indication of number of Non-Reserved connections on said link l, (i. e. : $N_{NRl}$).

Given these pieces of informations, not only the access node (SW0) Control Point is made capable of defining whether Non-Reserved traffic from DTE_A may be sent over the network, but in addition the network organization makes it possible to fairly and dynamically adjust the bandwidths distributed among the link connected Non-Reserved sources in order to mimize the data packets lost, without impacting on Reserved traffic.

Represented in figure 2 is a block diagram showing the various devices used within any given node, e. g. Node zero including a switching device SW0 and attached Control point device CP0. The switching device SW0 includes several adapters each being connected to one line or link said adapters are labelled Adapter 0, Adapter 1, ..., Adapter n. One line may handle up to thousands of connections to traffic sources. For example, traffic sources labelled 21, 22, 23 are represented attached to Adapter 0.

The data traffic of each source is oriented toward an Access Control Function device (see 24, 25, 26). Said Access Control Function devices are attached to an Adapter Control Agent (CA) in both directions. Each Control Agent (CA0, CA1, ... CAn) is attached to the node Control Point (CP0) to get, on request, link information for local connections attached, from the corresponding Topology Data Base. Also, as already mentioned, the CP0 Topology Data Base gets the Topology Data Update (TDU) messages (e.g. TDU 11, TDU 12...) from the network links (see figure 1).

Represented in figure 3 is an illustration of the operation of an Access Control Function (ACF) device 30 and a Connection Agent (CA) Device 31. To understand the operation of this system, one should first remember that the data traffic from any user is organized into packets (or ATM cells herein considered as packets) the transmission of which, over a node output line, is regulated. Various regulating methods are known in the art. One of these methods utilizes a so-called Leaky Bucket mechanism. A leaky bucket mechanism is illustraed in figure 4. Basically in the mechanism shown in figure 4, the data packets to be transferred are first passed through an admission buffer or shift register arrangement 41. The transfer from admission buffer to the network line is regulated by a token pool. To that end, a "token" generator is made to generate tokens at a predefined rate and said tokens are stored into a token pool 42. Then, each data packet to be transferred from the admission buffer 41, to the network, shall request as many tokens as the data packet contains bytes . Should these tokens be available in the token pool 42, the considered data packet shall be passed to the network. Otherwise the data packet should wait until the number of tokens has been generated. Since the tokens shall be individually attached to the requesting data packets, the mechanism may be improved to enable discriminating among discardable and non discardable data. Therefore, at the leaky bucket level, processing discrimination between discardable packets and non-discardable packets shall be performed through duplication of the token pool. To illustrate this, non-discardable packets shall have been tagged with a so-called green tag and discardable packets shall have been tagged with a so-called red tag. Accordingly two token pools are then used, one for "green" tokens and one for "red" tokens. Both pools are fed-in independently from each other. The green token pool is fed with a token rate $C\_eq^{(k)}$ equal to the equivalent capacity that has been reserved for the connection k in the network, or with a token rate $MCR_k$ which is the minimum guarateed bandwidth for NR. The red token pool is fed with a token rate $R_{k,t}$. While the computation of C_eq is described in the above mentioned references, the adaptive computation of the $R_{k,t}$ is an object of the present invention.

For instance, discardable data packets may include voice data packets recoverable at receiving end through several known mechanisms, e. g. interpolation/extrapolation mechanisms. But what is more important for the present invention is that the data packets transfers through the leaky bucket mechanism may be regulated by controlling the token generation rates. This is why, turning back to figure 3, one may notice that the Leaky Bucket 32 of the Access Control Function device (24) for a data source k is attached to a variable rate token pool 33. Said token pool 33 is also provided with threshold reference indications (a low threshold reference $TH_L$ and a high treshold reference $TH_H$) to be

used for regulation purposes. The token pool level with respect to these thresholds shall indicate the measure of utilisation and define whether the token generation rate should be increased, decreased or kept even (i. e. not changed).

For every connection along the considered path the token generation rate $R_{k,t}$ is updated by a mechanism. For instance, the Token Rate computing mechanism 31 for connection k, shall update the token generation at time t by providing an updated token generation rate $R_{k,t}$. This is made possible first through the use of a parameter measuring the utilization on connection k (device 35) monitored by the token pool 33 and indicating whether the token generation rate should be increased, decreased or not changed, for instance, based on monitoring the token pool level with reference to $TH_H$ and $TH_L$.

The informations provided by all the token utilization devices, such as 35, attached to the individual local connections of the Access Control Function device 30 attached to the considered access link, are fed into an allocated bandwidth computing device 36 within the Connection Agent device 31. This computing device 36 keeps tracking the bandwidth currently allocated to each connection k of the considered port. The computing device 36, in turn, drives a rate updating device 37 controlling the token rate device 34. The computing device 36, is provided, upon request, with all required Explicit Rates ($ER_l$) and numbers of Non-Reserved connections ($N_{NRl}$) available in the Topology Data Base as updated by the various $TDU_s$) fed into the considered node Control Point (e. g. CP0) to enable computing the updated token generation rate $R_{l,k,t}$ to be assigned to the connection k ou link l at time t, and then drive the token rate generation 34.

In operation, the system gets periodically leaky bucket measures and is able to decide if a connection needs its rate to be increased (i) decreased (d), or kept even (i. e. unchanged) (e).

Let $L_k$ be the set of links along the path of connection k. To determine the rate allowed to the connection k of the local port (e. g. SW0 for DTE_A) one first needs to know the portion $B_{l,t}$ of the bandwidth allocated to all the connections of the port, for every link l belonging to $L_k$ along their path. $B_{l,t}$ is given by :

$$B_{l,t} = ER_l \times N_l$$

$$= \frac{(1 - (Rho)^{res}_{l,t})\, Cl - \sum_{k=1}^{N_{NRl}} MCR_k}{N_{NRl}} N_l$$

wherein :

$N_l$ is the number of Non-Reserved connections attached to the local port that have link l in their path ;

$ER_l$ is the explicit rate for link l ;

$(Rho)^{res}_{l,t}$ is the bandwidth ratio used by reserved traffic on link l at time t, statistically monitored on the network nodes considered.

$C_l$ is the link l speed ;

$N_{NRl}$ is the total number of Non-Reserved connections within the network that have link l on their path.

$MCR_k$ is the fraction of bandwidth "reserved" for Non Reserved traffic of connection k. (Eventhough this parameter $MCR_k$ should be null for Non-Reserved traffic, one may optionally reserve a fair minimum bandwidth anyway, for that traffic). In other words

$\sum_{k=1}^{N_{NR}} MCR_k$ would be the sum of the Minimum Cell Rates

would be the sum of the Minimum Cell Rates (MCR) (if any) of all Non-Reserved connections sharing link l.

Then, using the leaky bucket measurements for the local node at time t - 1, the new explicit rates $R_{k,t}$ for sources attached to this node should be computed for time t.

Using informations from the leaky bucket, the system can deduce the number $N^{(i)}_{l,t}$ of connections bottlenecked asking for more bandwidth, $N^{(d)}_{l,t}$ from connections which do not use all their bandwidth and therefore can release some of it, $N^{(e)}_{l,t}$ of connections which use the bandwidth they need and do not need more ($N_l = N^{(i)}_{l,t} + N^{(d)}_{l,t} + N^{(e)}_{l,t}$). Note that this classification based on whether the rate of connection k should be increased, decreased or unchanged only depends on the current value set at the local node, $R_{k,t-1}$.

For a given connection k, we will use the current rate $R_{l,k,t-1}$ set on link l at time t-1 and re-write $R_{l,k,t-1}$ as $R^{(d)}_{l,k,t-1}$ if connection k needs a decrease, $R^{(i)}_{l,k,t-1}$ if connection k needs an increase and $R^{(e)}_{l,k,t-1}$ if connection k can be held to its current rate. The computation of the rates $R_{l,k,t}$ is done according to the following recurrent system on every link l of $L_k$ :

Ø if decrease,

$$R_{l,k,t} = \frac{B_{l,t}}{B_{l,t-1}} [R^{(d)}_{l,k,t-1} \cdot \alpha'_{l,t}]$$

Ø if no change,

$$R_{l,k,t} = \frac{B_{l,t}}{B_{l,t-1}} [R^{(e)}_{l,k,t-1} \cdot \alpha''_{l,t} + \beta'_{l,t}]$$

Ø if increase,

$$R_{l,k,t} = \frac{B_{l,t}}{B_{l,t-1}} \left[ \frac{1}{N^{(i)}_{l,t}} \sum_{j=1}^{N^{(i)}_{l,t}} R^{(i)}_{l,j,t-1} + \beta''_{l,t} \right]$$

where $\alpha'_{l,t}$ and $\alpha''_{l,t}$ are multiplicative decrease factors and $\beta'_{l,t}$ and $\beta''_{l,t}$ are additive increase factors.

The rate of connection k is then set to the minimum rate computed along its path :

$$R_{k,t} = \min_{l \in L_k} \{R_{l,k,t}\}$$

and then,

$$R_{k,t} = \max \{MCR_k, \min \{PCR_k, R_{k,t}\}\}$$

The underlying assumption for this model is that all connections should be able to release at any time the bandwidth they got in order to satisfy a defined fairness criterion. For instance, if at an arbitrary time there are only connections which ask for an increase and connections which are "ok", it is not fair that this state for the previous system be stable : some of the bandwidth allocated to the connections "ok" should be re-distributed among the bottlenecked connections. The fairness criterion chosen here means that a state where all connections need more bandwidth than the network can offer is a more fair state than the unbalanced state mentioned here above. This is, naturally, but one of the fairness criterion which could be used.

The sum term in the equation for connections which need more bandwidth is also used to achieve some fairness among them. Without this term (i. e., simply using $R^{(i)}_{l,k,t-1}$), connections which already have a significant amount of the total bandwidth and ask for more will never release part of what they got in favor of connections with small rates which suddenly ask for bandwidth. As this scenario is not fair, the global bandwidth of connections waiting for an increase should be equally redistributed (according to the max-min criterion).

At last, the term $B_{l,t}/B_{l,t-1}$ equally distributes over all connections the variations of the available bandwidth for non reserved traffic on link l.

The problem for computing $\alpha'_{l,t}$, $\alpha''_{l,t}$, $\beta'_{l,t}$ and $\beta''_{l,t}$ is not simple, as they are function of one or more of the time variables $N^{(i)}_{l,t}$, $N^{(d)}_{l,t}$, $N^{(e)}_{l,t}$. As a consequence, they are themselves function of time. Moreover, these parameters should always satisfy, for link l, that

$$\sum_{j=1}^{N_l} R_{l,j,t} = B_{l,t}/B_{l,t-1} \times \sum_{j=1}^{N_l} R_{l,j,t-1}$$

The resolution of the system being too complex to be evaluated for each time t, an heuristic approach should be employed :

Ø For connections which should decrease, $\alpha'_{l,t}$ is of course less than one and should be closer to 0 as the number of connections which need bandwidth is high. Some of the bandwidth released should also be added to connections which need no change, i. e., is part of the $\beta'_{l,t}$ term. We set :

$$\alpha'_{l,t} = 1 - \frac{N^{(i)2}_{l,t}}{N^2_l} - \frac{N^{(i)}_{l,t} N^{(e)}_{l,t}}{N^2_l} \quad (1)$$

Ø For connections which should need equal bandwidth, $\alpha''_{l,t}$ should also be closer to 0 as the number of connections which need bandwidth is high.

$$\alpha''_{l,t} = 1 - \frac{N^{(i)2}_{l,t}}{N^2_l} \quad (2)$$

On the other hand, the rate of these connections grows up by applying $N^{(i)}_{l,t} N^{(e)}_{l,t}/N^2_l$ on connections which release bandwidth.

This amount is equally distributed over the $N^{(e)}_{lt}$ connections ; this gives :

$$\beta'_{l,t} = \frac{N^{(i)}_{l,t}}{N^2_l} \sum_{j=1}^{N^{(e)}_{l,t}} R^{(d)}_{l,j,t-1} \quad (3)$$

Ø For connections which should increase, the term $\beta''_{l,t}$ is the total bandwidth released by applying the term $N^{(i)2}_{l,t}/N^2_l$ on both previous sets of connection rates, equally distributed over the $N^{(i)}_{l,t}$ connections :

$$\beta''_{l,t} = \frac{N^{(i)}_{l,t}}{N^2_l} \left( \sum_{j=1}^{N^{(d)}_{l,t}} R^{(d)}_{l,t-1} + \sum_{j=1}^{N^{(e)}_{l,t}} R^{(e)}_{l,j,t-1} \right) \quad (4)$$

The algorithm used to implement the invention may be summarized as follows :

* For all connections :

Ø using the leaky bucket statistics, "mark" connection k as requiring an increase, decrease, or no change.

* For all links :

Ø Compute $N^{(i)}_{l,t}$, $N^{(d)}_{l,t}$ and $N^{(e)}_{l,t}$

Ø Upon reception of a TDU message, update the available bandwidth available on each link l:

$$B_{l,t} = ER_l \times N_l = \frac{(1 - (Rho)^{res}_{l,t})C_l - \sum_{k=1}^{N_{NRI}} MCR_k}{N_{NRI}} N_l \quad (5)$$

Ø Compute $\sum_{j=1}^{N^{(d)}_{l,t}} R^{(d)}_{l,j,t-1}$, $\sum_{j=1}^{N^{(i)}_{l,t}} R^{(i)}_{l,j,t-1}$ and $\sum_{j=1}^{N^{(e)}_{l,t}} R^{(e)}_{l,j,t-1}$

* For all connections :

Ø For all links in connection path, update $R_{l,k,t}$ :

- if decrease,

$$R_{l,k,t} = \frac{B_{l,t}}{B_{l,t-1}}[R^{(d)}_{l,k,t-1}(1 - \frac{N^{(i)2}_{l,t}}{N^2_l} - \frac{N^{(i)}_{l,t}N^{(e)}_{l,t}}{N^2_l})] \quad (6)$$

- if no change,

$$R_{l,k,t} = \frac{B_{l,t}}{B_{l,t-1}} = [R^{(e)}_{l,k,t-1}(1 - \frac{N^{(i)2}_{l,t}}{N^2_l}) + \frac{N^{(i)}_{l,t}}{N^2_l} \times \sum_{j=1}^{N^{(d)}_{l,t}} R^{(d)}_{l,j,t-1}] \quad (7)$$

- if increase,

$$R_{l,k,t} = \frac{B_{l,t}}{B_{l,t-1}} [\frac{1}{N^{(i)}_{l,t}} \sum_{j=1}^{N^{(i)}_{l,t}} R^{(i)}_{l,j,t-1} + \frac{N^{(i)}_{l,t}}{N^2_l} (\sum_{j=1}^{N^{(d)}_{l,t}} R^{(d)}_{l,j,t-1} + \sum_{j=1}^{N^{(e)}_{l,t}} R^{(e)}_{l,j,t-1})] \quad (8)$$

Ø and set

$$R_{k,t} = \min_{l \in L_k} \{R_{l,k,t}\} \quad (9)$$

$$R_{k,t} = \max\{MCR_k, \min\{PCR_k, R_{k,t}\}\} \quad (10)$$

In practice, it is expected that the non-reserved traffic will be mostly bursty with long idle periods. Because of these silent phases, it is reasonable to think that only a few connections will be active and then performances would be increased by introducing some statistical multiplexing. Basically, this means that we could allocate to every connection a "little bit more" than its rough fair share, as long as there is no congestion within the network.

Represented in figure 5 is a block diagram of the algorithm for implementing the invention and getting the rate adjusment looked for, for a connection k.

First, a measure of utilization as defined by device 35 (see figure 3), as well as the measures of utilization of all other local connections (1, 2, 3, ...k-1) is provided to a computing stage 50. Said computing stage also gets the rates at t-1 from a TABLE stored in the Connection Agent (CA) considered. Said table looks as follows :

TABLE

| | link 1 | link 2 | -- | link 1 |
|---|---|---|---|---|
| Connection 1 | $R_{1,1,t-1}$ | -- | -- | $R_{l,1,t-1}$ |
| Connection 2 | - | - | - | - |
| - | - | - | - | - |
| Connection k | - | - | - | $R_{l,k,t-1}$ |

The computing stage 50 (see figure 5) provides then $N^{(i)}_{l,t}$, $N^{(d)}_{l,t}$ and $N^{(e)}_{l,t}$.

These informations, together with the data provided by the above TABLE are then passed through a second computing stage 51, wherein the parameters $\alpha'_{l,t}$, $\alpha''_{l,t}$, $\beta'_{l,t}$ and $\beta''_{l,t}$ are computed according to the above mentioned equations 1 through 4, respectively.

Finally, the $\alpha$ and $\beta$ parameters, are fed into a third computing stage 52, together with the Explicit Rates for link l and the number of Non-Reserved traffics $N_{NRl}$, to enable computing the updated available bandwidth $B_{l,t}$ according to equation 5.

The above computing operations of stages 50, 51 and 52 are performed for every link l.

Then, for every connection along the considered path, and depending whether the state of the connection as defined by the measure of utilization is: decrease, no change, (i. e. keep even), or increase, then the link rate is updated according to equations 6, 7 or 8 respectively, as indicated in stages 53, 54 and 55 of figure 5 (and device 37 in figure 3).

Finally, the new token rate for link k (see device 34 in figure 3), is computed in stage 56 through computation of equations (9) and (10).

This new token generation rate $R_{k,t}$ is then applied to update the token generation rate for the token pool 33 (see figure 3).

## Claims

1. An adaptive bandwidth allocation method for optimizing data traffic in a high speed data transmission network including Nodes interconnected by high speed links made for vehiculating data traffics of different priority levels along assigned network paths between data terminal equipments acting as data sources and data destination terminals, said priority levels including high priority level(s) for so-called Reserved traffic for which a transmission bandwidth has been reserved along said path based on predefined agreements, and low priority level for so-called Non-Reserved traffic which should be transferred over the network with the transmission bandwidth available along the considered path once Reserved traffic is satisfied, said method for assigning bandwidth to Non-Reserved traffic including :

   - generating at least one Topology Data Base storing image of the network occupancy over each link along the network paths;

   - periodically and at any terminal call set-up within the network, generating and broadcasting Topology Data Base updating (TDUs) messages and storing in said at least one Topology Data Base said TDUs, said TDU messages including a so-called Explicit Rate parameter ($ER_l$) for each link l indicating the bandwidth currently

available on link l, and a parameter $N_{NR}$ indicating the number of Non-Reserved connections on link l;

- upon receiving said TDUs information, computing for each node along the considered path the amount of transmission bandwidth left available over each link along the considered path and assigning said available bandwidth to the Non-Reserved traffic sources connected to the network.

2. A adaptive bandwidth allocation method according to claim 1 wherein said transmission network is a distributed control network said Topology Data Base being stored into each network node and said TDUs being broadcasted to said nodes along the considered path, said bandwidth available for Non-Reserved traffic being computed and assigned to concerned links within each node along the path.

3. An adaptive bandwidth allocation method according to claims 1 or 2 including the following steps :

a - for all connections k over a link l, mark periodically and at any connection set-up, the connections as requiring an increased (i) bandwidth or a decreased (d) bandwidth, or no change (e) ;

b - for all links l attached to a node along the considered path

- compute the number of connections requesting more bandwidth ($N_{l,t}^{(i)}$), the number of connections requiring less bandwidth ($N_{l,t}^{(d)}$) and the number of connections owning the amount of bandwidth they need ($N_{l,t}^{(e)}$) ;

- upon reception of a TDU message, update the available bandwidth :

$$B_{l,t} = \frac{(1 - (Rho)_{l,t}^{res})C_l - \sum_{k=1}^{N_{NRl}} MCR_k}{N_{NRl}} N_l$$

wherein

$N_l$ is the number of Non-Reserved connections attached to the local ports that have link l in their path ;

$(Rho)^{res}_{l,t}$ is the bandwidth ratio for reserved traffic on link l at time t, statistically monitored on the network nodes considered.

$C_l$ is the link l speed ;

$N_{NRl}$ is the total number of Non-Reserved connections within the network that have link l on their path.

$MCR_k$ is the fraction of bandwidth "reserved" for Non Reserved traffic of connection k. (Eventhough this parameter $MCR_k$ should be null for Non-Reserved traffic, one may optionally reserve a fair minimum bandwidth anyway, for that traffic). In other words

$\sum_{k=1}^{N_{NR}} MCR_k$ would be the sum of the Minimum Cell Rates

- Compute

$$\sum_{j=1}^{N^{(d)}_{l,t}} R^{(d)}_{l,j,t-1}, \sum_{j=1}^{N^{(i)}_{l,t}} R^{(i)}_{l,j,t-1} \text{ and}$$

$$\sum_{j=1}^{N^{(e)}_{l,t}} R^{(e)}_{l,j,t-1}$$

c - for all connections and for all links in connection path update $R_{l,k,t}$ as follows :

- if decrease required :

$$R_{l,k,t} = \frac{B_{l,t}}{B_{l,t-1}} \left[ R^{(d)}_{l,k,t-1} \left(1 - \frac{N^{(i)2}_{l,t}}{N^2_l} - \frac{N^{(i)}_{l,t} N^{(e)}_{l,t}}{N^2_l}\right)\right]$$

- if no change required :

$$R_{l,k,t} = \frac{B_{l,t}}{B_{l,t-1}} = \left[R^{(e)}_{l,k,t-1} \left(1 - \frac{N^{(i)2}_{l,t}}{N^2_l}\right) + \frac{N^{(i)}_{l,t}}{N^2 l} \times \sum_{j=1}^{N^{(d)}_{l,t}} R^{(d)}_{l,j,t-1}\right]$$

- if increase required :

$$R_{l,k,t} = \frac{B_{l,t}}{B_{l,t-1}} \left[ \frac{1}{N^{(i)}_{l,t}} \sum_{j=1}^{N^{(i)}_{l,t}} R^{(i)}_{l,j,t-1} + \frac{N^{(i)}_{l,t}}{N^2_l} \left( \sum_{j=1}^{N^{(d)}_{l,t}} R^{(d)}_{l,j,t-1} + \sum_{j=1}^{N^{(e)}_{l,t}} R^{(e)}_{l,j,t-1} \right) \right]$$

and finally setting $R_{k,t} = \min \{ R_{l,k,t} \}$

4. A system for dynamically adapting transmission bandwidth allocation to data traffic over a high speed transmission network including switching nodes (SW0, SW1....) attached to at least one Control Point processing unit keeping at least one updated Topology Data Base representing an instaneous image of the network, said switching nodes being interconnected by high speed links (10, 11...) for vehiculating data traffic of different priority levels along assigned network paths between network connected data sources and destination terminals, said priority levels including a high priority level for so-called Reserved traffic for which a transmission bandwidth is reserved in each link along the path based on predefined conditions, and a low priority level for so-called Non-Reserved traffic which should be transfered over the network within the transmission bandwidth left available along the considered path once Reserved traffic is optimized, said system for dynamically assigning transmission bandwidth to Non-Reserved traffic including :

- an Access Control Function device including Access Control means (24) connected to each data traffic source (k) attached to the considered link "l" for monitoring the source traffic and measuring current utilization from said connection k;

- a Connection Agent (CA) (31) connected to each Control Function device of a data traffic source attached to a same link "l" ;

- means for attaching said Control Agent device to said at least one Topology Data Base to get, therefrom, link informations on request ;

- first computing means (36) within said Connection Agent device for computing the portion of line bandwidth $B_{l,t-1}$ currently allocated to each connection k on the considered link l ;

- second computing (37, 34) for computing an updated rate $R_{k,t}$ assignable to each Non-Reserved traffic source of each connection k attached to link l, through said Access Control Function means (24).

5. A system for dynamically adapting transmission bandwidth allocation according to claim 3 in a distributed network system wherin each network node is assigned a Control Point device (CP) storing said Topology Data Base, said system being provide with means for broadcasting Topology Data base Updating messages (TDUs) including, for each considered link l, a so-called Explicit Rate (ERl) parameter indicating the current available bandwidth on link l, and a parameter $N_{NR}$ indicating the number of Non-Reserved connections currently attached to said link l.

6. A system for dynamically adapting transmission bandwidth allocation, as claimed in claims 4 or 5 wherein said Control Function means (24) includes :

- a leaky bucket (32) connected to each data source connection k for receiving data packets therefrom and to be passed to next link on the considered path upon collecting a so-called token from a token pool (33) filled in at a token generation rate $R_{k,t}$ ;

- means (35) for periodically monitoring the token pool (33) content with respect to predefined threshold levels and for measuring the considered connection k requirements based on the token pool level with respect to said thresholds.

7. A system for dynamically adapting transmission bandwidth allocation as claimed in claim 6, wherein said token pool is duplicated to enable distinguishing between Reserved traffic and Non-Reserved traffic.

8. A system for dynamically adapting transmission bandwidth allocation according anyone of claims 6 or 7 wherein said means (35) for measuring each connection utilization periodically, monitors the corresponding token pool level of filling with respect to a so-called low threshold ($TH_L$) and a so-called high threshold ($TH_H$) and generate accordingly an indication meaning whether the token pool filling rate should be increased (i), decreased (d), or not changed (e).

9. A system for dynamically adapting transmission bandwidth allocation as claimed in claim 8, wherein said Control Agent device (31) includes :

- means (36) for measuring the portion $B_{l,t}$ of bandwidth allocated to all the connections for every link l,

$$B_{l,t} = ER_l \times N_l$$

wherein = $ER_l$ is the Explicit Rate for link l
$N_l$ is the number of Non-Reserved connections attached to the nodes that have link l on their path,

- rate updating means (37) connected to said $B_{l,t}$ measuring means (36) and to said measure of utilization means (35) for computing the rate to be assigned to packet transfers on every connection k over link l at time t, based on previous value ($R_{l,kt-1}$) at time t-1 and increase/decrease/no change indications provided by means (35)

whereby :
for decrease :

$$R_{l,k,t} = \frac{B_{l,t}}{B_{l,t-1}} [R^{(d)}{}_{l,k,t-1} \cdot \alpha'_{l,t}]$$

for no change :

$$R_{l,k,t} = \frac{B_{l,t}}{B_{l,t-1}} [R^{(e)}{}_{l,k,t-1} \cdot \alpha''_{l,t} + \beta'_{l,t}]$$

for increase :

$$R_{l,k,t} = \frac{B_{l,t}}{B_{l,t-1}} \left[ \frac{1}{N^{(i)}_{l,t}} \sum_{j=1}^{N^{(i)}_{l,t}} R^{(i)}_{l,j,t-1} + \beta''_{l,t} \right]$$

wherein : $\alpha'$ and $\alpha''$ are predefined multiplicative decrease factors
$\beta'$ and $\beta''$ are predefined additive increase factors,

wherein the indexes (d), (e), (i) respectively identify the connections that can release some of their bandwidths, those which own the bandwidth they need, and those asking for more bandwidth as indicated by devise (35) ;

- means (34) for setting the updated token generation rate for connection k to the minimum rate along its path :

$$R_{k,t} = \min \{R_{l,k,t}\}$$

**10.** A system for dynamically adapting transmission bandwidth allocation as claimed in claim 9. Wherein the minimum fraction of bandwidth $MCR_k$ reserved for Non-Reserved connection is made equal to zero.

**11.** A system for dynamically adapting transmission bandwidth allocation as claimed in claim 9 wherein $R_{k,t}$ is bound by $MCR_k$ and $PCR_k$, with :

$$R_{k,t} = \max \{MCR_k, \min \{PCR_k, R_{k,t-1}\}\}$$

**12.** A system for dynamically adapting transmission bandwidth allocation as claimed in anyone of claims 9, 10 or 11, wherein:

$$\alpha'_{l,t} = 1 - \frac{N^{(i)2}_{l,t}}{N^2_l} - \frac{N^{(i)}_{l,t} N^{(e)}_{l,t}}{N^2_l}$$

$$\alpha''_{l,t} = 1 - \frac{N^{(i)2}_{l,t}}{N^2_l}$$

$$\beta'_{l,t} = \frac{N^{(i)}_{l,t}}{N^2_l} \sum_{j=1}^{N^{(d)}_{l,t}} R^{(d)}_{l,j,t-1}$$

$$\beta''_{l,t} = \frac{N^{(i)}_{l,t}}{N^2_l} \left( \sum_{j=1}^{N^{(d)}_{l,t}} R^{(d)}_{l,t-1} + \sum_{j=1}^{N^{(e)}_{l,t}} R^{(e)}_{l,j,t-1} \right)$$

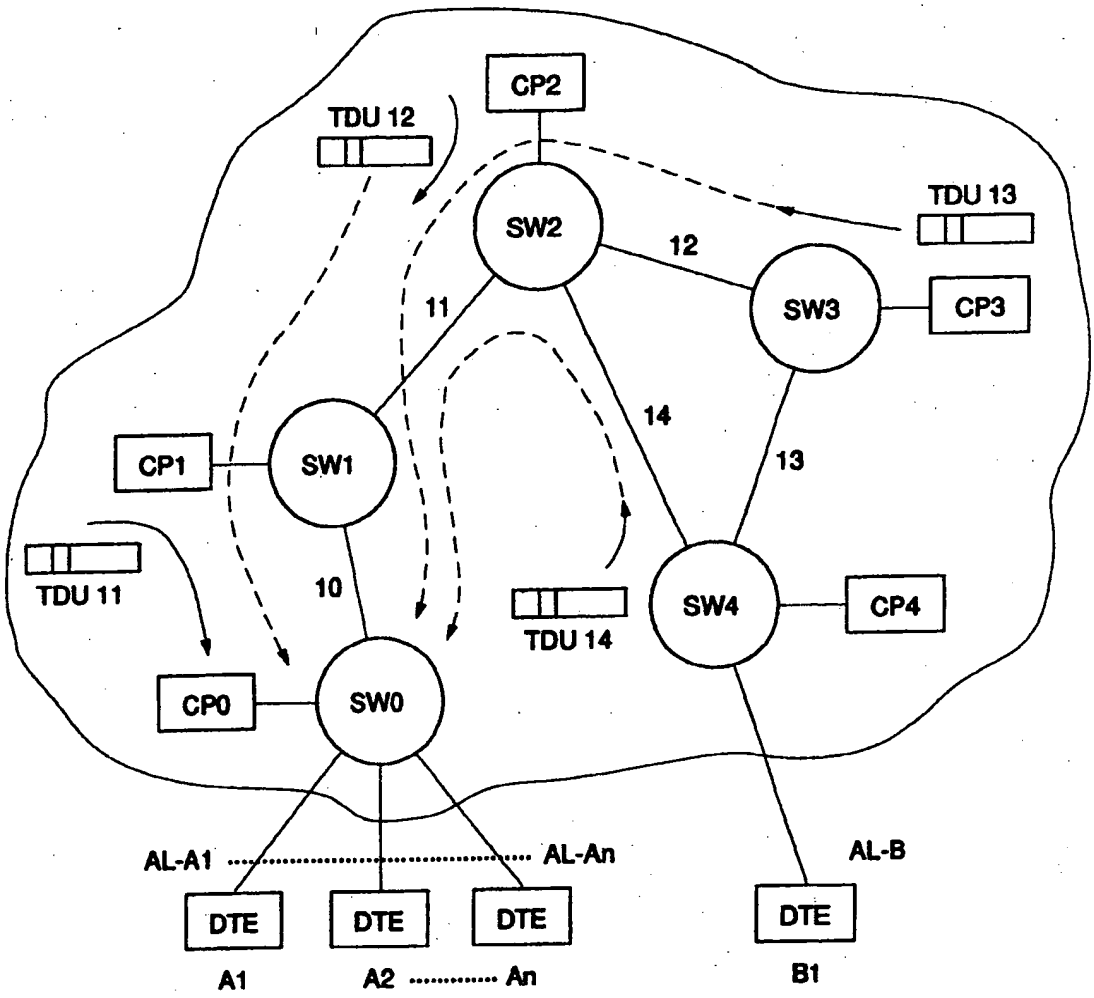
CP2
TDU 12
TDU 13
SW2
12
SW3
CP3
11
CP1
SW1
14
13
TDU 11
10
TDU 14
SW4
CP4
CP0
SW0
AL-A1
AL-An
AL-B
DTE
DTE
DTE
DTE
A1
A2
An
B1

FIG.1

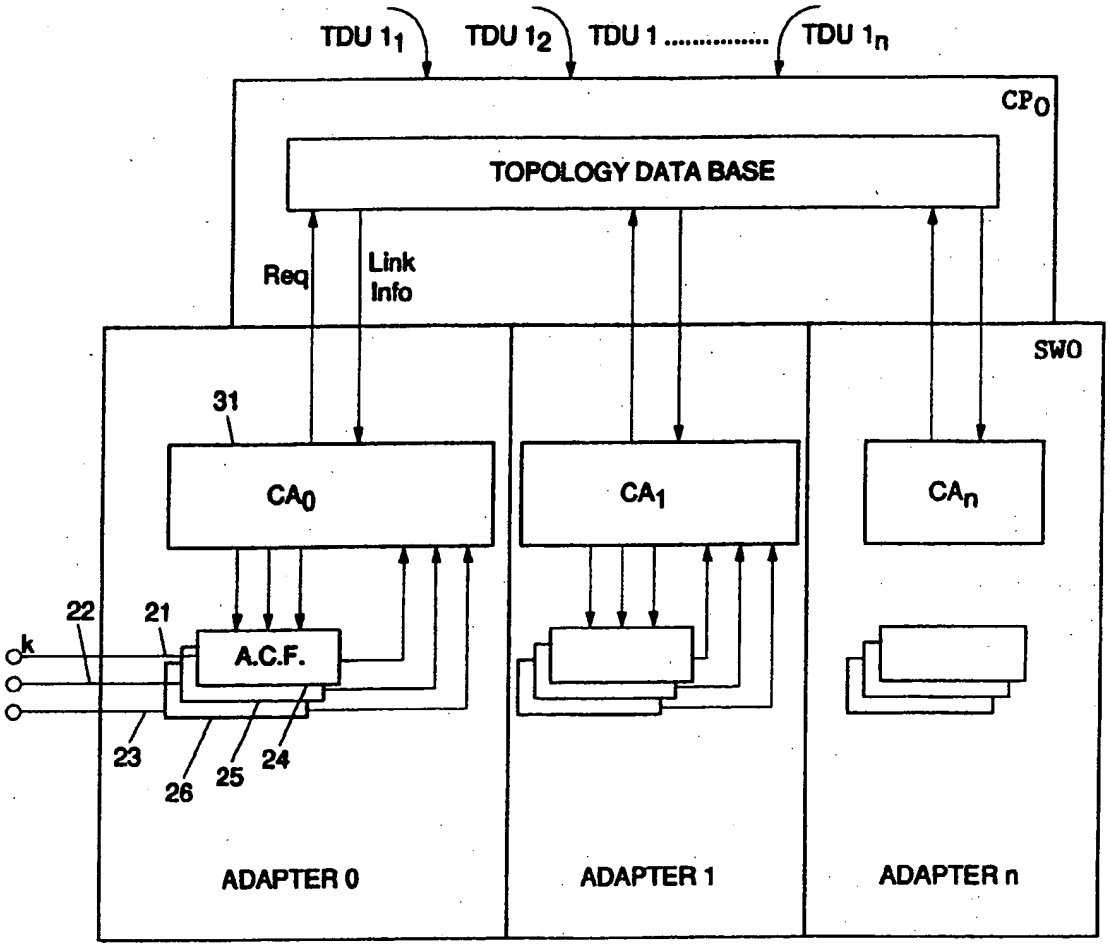
TDU 1$_1$
TDU 1$_2$
TDU 1 .....
TDU 1$_n$
CP$_0$
TOPOLOGY DATA BASE
Req
Link Info
SWO
31
CA$_0$
CA$_1$
CA$_n$
22
21
k
A.C.F.
23
26
25
24
ADAPTER 0
ADAPTER 1
ADAPTER n

FIG.2

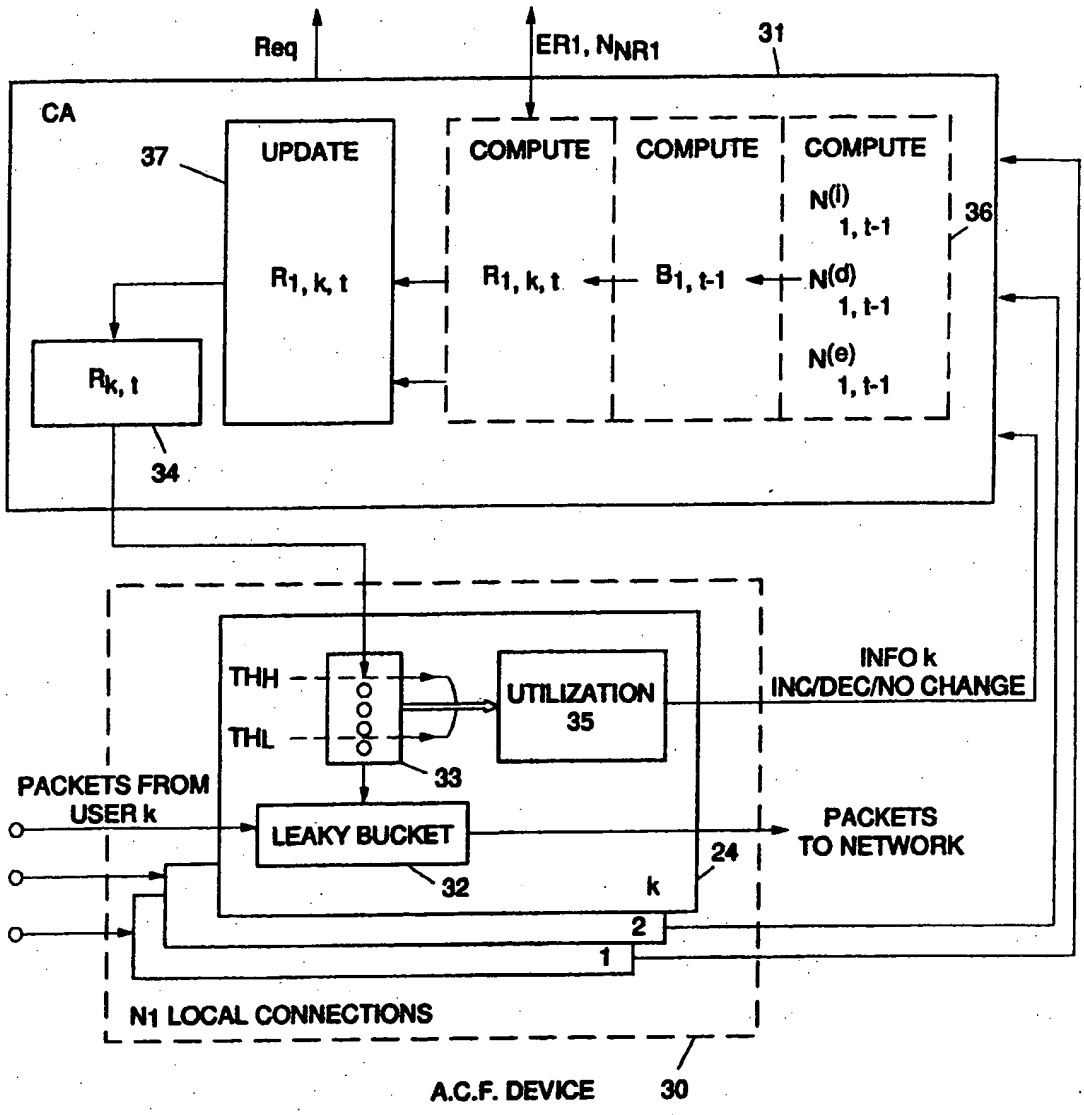
Req
ER1, $N_{NR1}$
31
CA
37
UPDATE
COMPUTE
COMPUTE
COMPUTE
$N^{(i)}_{1,t-1}$
36
$R_{1,k,t}$
$R_{1,k,t}$
$B_{1,t-1}$
$N^{(d)}_{1,t-1}$
$N^{(e)}_{1,t-1}$
$R_{k,t}$
34
THH
THL
33
UTILIZATION
35
INFO k
INC/DEC/NO CHANGE
PACKETS FROM
USER k
LEAKY BUCKET
32
PACKETS
TO NETWORK
24
k
2
1
N1 LOCAL CONNECTIONS
A.C.F. DEVICE
30

FIG.3

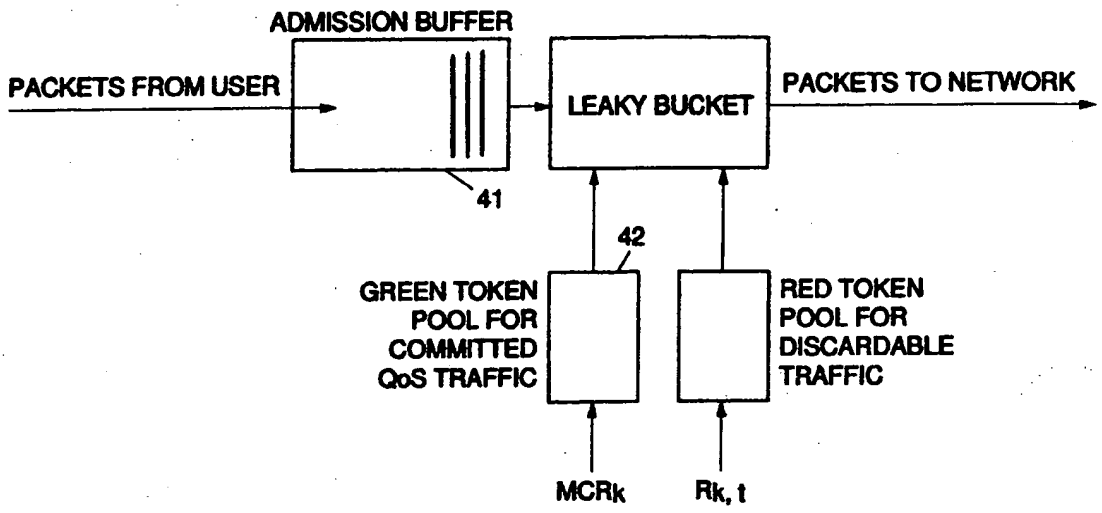
ADMISSION BUFFER
PACKETS FROM USER
LEAKY BUCKET
PACKETS TO NETWORK
41
42
GREEN TOKEN POOL FOR COMMITTED QoS TRAFFIC
RED TOKEN POOL FOR DISCARDABLE TRAFFIC
MCRk
Rk, t

FIG.4

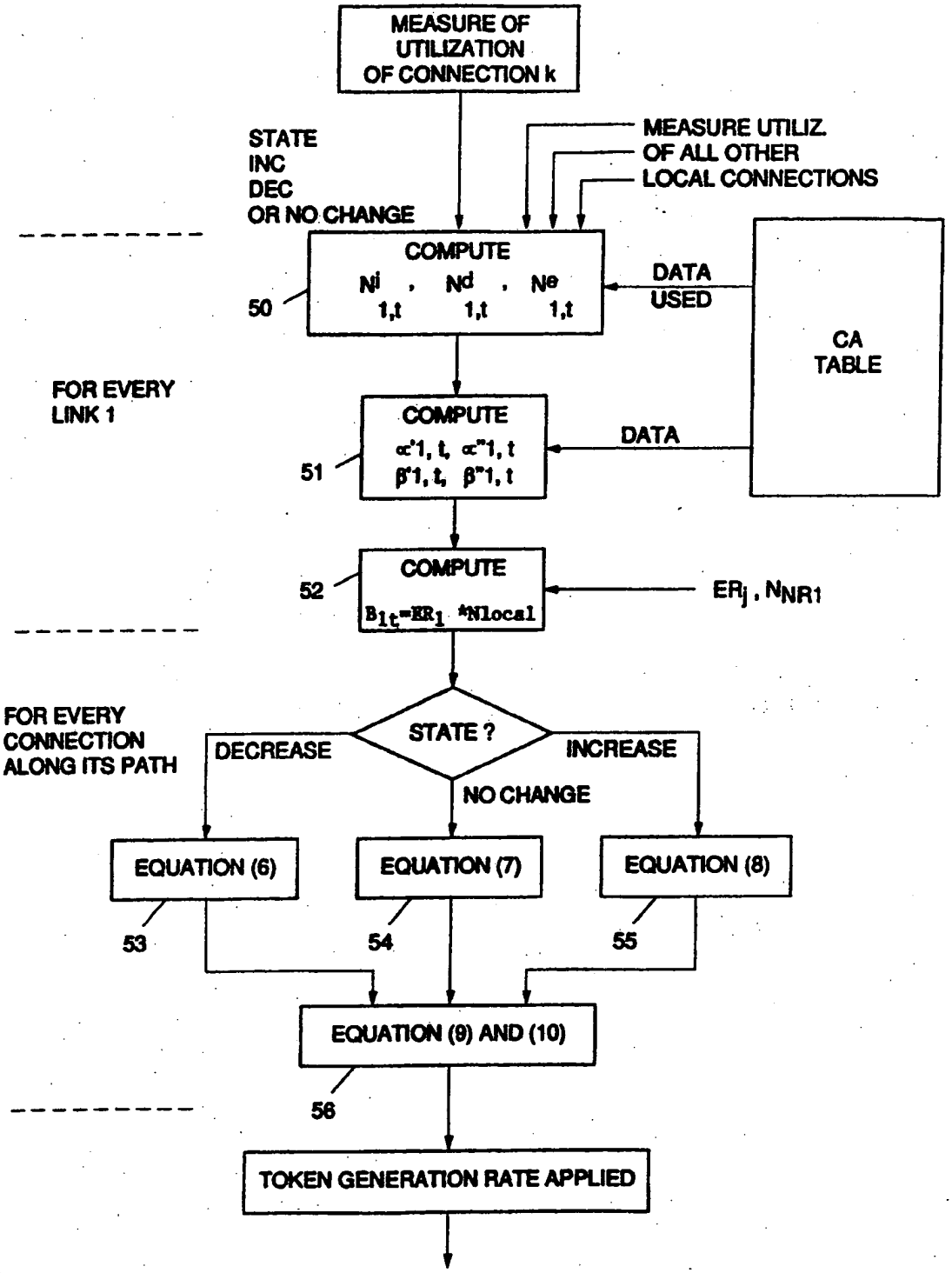
MEASURE OF UTILIZATION OF CONNECTION k
STATE INC DEC OR NO CHANGE
MEASURE UTILIZ. OF ALL OTHER LOCAL CONNECTIONS
COMPUTE
50
DATA USED
CA TABLE
FOR EVERY LINK 1
COMPUTE
51
DATA
COMPUTE
52
ERj , NNR1
FOR EVERY CONNECTION ALONG ITS PATH
STATE ?
DECREASE
INCREASE
NO CHANGE
EQUATION (6)
EQUATION (7)
EQUATION (8)
53
54
55
EQUATION (9) AND (10)
56
TOKEN GENERATION RATE APPLIED

FIG.5

European Patent Office

# EUROPEAN SEARCH REPORT

Application Number
EP 95 48 0182

## DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document with indication, where appropriate, of relevant passages | Relevant to claim | CLASSIFICATION OF THE APPLICATION (Int.Cl.6) |
|---|---|---|---|
| Y | US-A-5 347 511 (GUN LEVENT) 13 September 1994<br>* column 1, line 21 *<br>* line 32 - line 34 *<br>* column 2, line 1 - line 10 *<br>* column 2, line 46 *<br>* column 4, line 5 - line 9 *<br>* column 7, line 27 - line 32; figure 4 *<br>--- | 1,2,4 | H04Q11/04<br>H04L12/56 |
| Y | NEW ORLEANS SUPERCOMM/ICC '94. SERVING HUMANITY THROUGH COMMUNICATIONS. 1994 IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS. CONFERENCE RECORD (CAT. NO.94CH3403-3), PROCEEDINGS OF ICC/SUPERCOMM'94 - 1994 INTERNATIONAL CONFERENCE ON COMMUNICATIONS, N, ISBN 0-7803-1825-0, 1994, NEW YORK, NY, USA, IEEE, USA,<br>pages 1592-1599 vol.3, XP000438764<br>BAHK S ET AL: "Preventive congestion control based routing in ATM networks"<br>* page 1593, left-hand column, line 45 - right-hand column, line 63 *<br>---<br>-/-- | 1,2,4 | |

TECHNICAL FIELDS SEARCHED (Int.Cl.6)

H04L
H04Q

The present search report has been drawn up for all claims

| Place of search | Date of completion of the search | Examiner |
|---|---|---|
| THE HAGUE | 24 May 1996 | Veen, G |

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

European Patent Office

# EUROPEAN SEARCH REPORT

Application Number
EP 95 48 0182

## DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document with indication, where appropriate, of relevant passages | Relevant to claim | CLASSIFICATION OF THE APPLICATION (Int.Cl.6) |
|---|---|---|---|
| A | IEEE INFOCOM ' 93. THE CONFERENCE ON COMPUTER COMMUNICATIONS PROCEEDINGS. TWELFTH ANNUAL JOINT CONFERENCE OF THE IEEE COMPUTER AND COMMUNICATIONS SOCIETIES. NETWORKING: FOUNDATION FOR THE FUTURE (CAT. NO.93CH3264-9), SAN FRANCISCO, CA, USA, 28 MARCH-, ISBN 0-8186-3580-0, 1993, LOS ALMITOS, CA, USA, IEEE COMNPUT. SOC. PRESS, USA, pages 376-383 vol.1, XP000419753 HARTANTO V F ET AL: "User-network policer: a new approach for ATM congestion control" * page 379, left-hand column, line 17 - right-hand column, line 28; figure 3 * ----- | 6,7 | |

TECHNICAL FIELDS SEARCHED (Int.Cl.6)

The present search report has been drawn up for all claims

| Place of search | Date of completion of the search | Examiner |
|---|---|---|
| THE HAGUE | 24 May 1996 | Veen, G |

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)